International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:1, No:12, 2007

Facial Expressions Animation and Lip Tracking Using Facial Characteristic Points and Deformable Model

Hadi Seyedarabi, Ali Aghagolzadeh, and Sohrab Khanmohammadi

***Abstract*—**Face and facial expressions play essential roles in interpersonal communication. Most of the current works on the facial expression recognition attempt to recognize a small set of the prototypic expressions such as happy, surprise, anger, sad, disgust and fear. However the most of the human emotions are communicated by changes in one or two of discrete features. In this paper, we develop a facial expressions synthesis system, based on the facial characteristic points (FCP's) tracking in the frontal image sequences. Selected FCP's are automatically tracked using a cross-correlation based optical flow. The proposed synthesis system uses a simple deformable facial features model with a few set of control points that can be tracked in original facial image sequences.

***Keywords*—**Deformable face model, facial animation, facial characteristic points, optical flow.

I. Introduction

FACE expression recognition is useful for designing the new interactive devices offering the possibility of new ways for human to interact with the computer systems. Automating facial expression analysis and synthesis could bring facial expressions into man-machine interaction. Modeling and animation of the individualized face models is a very important and challenging problem in the computer graphics.

In 1971 Ekman and Friesen [1] postulated six primary emotions that each posses a distinctive content together with a unique facial expression. These prototypic emotional displays are also referred to as basic emotions. They seem to be universal across the human ethnicities and cultures and comprise happiness, sadness, fear, disgust, surprise, and anger. In the recent years, there has been a great amount of research on the face recognition, facial expression recognition and facial animation.

Zhilin Wu et al. [2] have presented a combined method to accurately track the outer lips by using the Gradient Vector Flow (GVF) snakes with parabolic templates as an additional external force. This combination needs fewer requirements of both salient boundaries and accuracy of templates.

Hadi Seyedarabi, Ali Aghagolzadeh, and Sohrab Khanmohammadi are with Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran {seyedarabi, aghagol, khan}@tabrizu.ac.ir

Furthermore, it is more flexible in tracking the noisy signals. They have also presented an inner lip tracking method using a similarity function with the temporal smoothing. An encoder to generate MPEG-4 facial animation parameters (FAPs) from the continuous lip boundaries is also designed

Eisert et al. [3] showed that the traditional waveform coding and 3-D model-based coding are not competing alternatives, but should be combined to support and complement each other. Both approaches are combined such that the generality of the waveform coding and the efficiency of 3-D model-based coding are available where needed. The combination is achieved by providing the block-based video coder with a second reference frame for prediction, which is synthesized by the model-based coder. The model-based coder uses a parameterized 3-D head model, specifying shape and color of a person. They restricted their investigations to the typical videotelephony scenarios that showed head-and-shoulder scenes. Motion and deformation of the 3-D head model constitute facial expressions which are represented by FAP's based on the MPEG-4 standard. An intensity gradient-based approach that exploits the 3-D model information is used to estimate the FAP's, as well as the illumination parameters, that describe changes of the brightness in the scene

2-D or 3-D facial model coding can be employed in various mobile applications to provide an enhanced user experience. The potential for the low bit rate and the scalability that 2-D or 3-D coding provides, makes it suitable for using in conjunction with the various mobile networks. Worral et al. [4] examined the performance of the low bit rate 3-D "Talking Heads" in order to determine the appropriate delivery scheme for the encoded data.

Erol [5] described a facial modeling and animation system that used muscle-based generic face model and deformed it using deformation techniques to model the individualized faces. Two orthogonal photos of the real faces were used for this purpose. Image processing techniques were employed to extract certain features on the photographs, which were then refined manually by the user through the facilities of the user interface of the system. The feature points located on the frontal and side views of a real face were used to deform the generic model. Then, the muscle vectors in the individualized face model were arranged accordingly. The individualized face models produced in this manner were animated using the parametric interpolation techniques.

In this paper we develop a deformable muscle-based face model that tracks some FCP's in the real face image sequences and shows the same expressions. Proposed deformable model has a simple structure and uses a few set of control points comparing to the similar face models. Proposed model has a low complexity and is suitable for the real time implementations, such as real time facial animation and video conference systems.

II. Image Database

In this work, we used Cohn-Kanade database [6] that consists of expression sequences of subjects, starting from a neutral expression and ending in the peak of the facial expression. Also we used a database that prepared at University of Tabriz for this research. There are 104 subjects in the Cohn-Kanade database. Subjects sat directly in front of the camera and performed a series of the facial expressions that included the six primary emotions, single Action Unit (AU), (e.g. AU25) and combinations of AUs (e.g. AU6+12+25). Since for subjects, not all of the primary emotions, single and composite AUs sequences were available to us, we used a subset of subjects. For each person there are on average 12 frames for each expression. Image sequences for the frontal views were digitized into 640×490 pixel array with 8 bits grayscale. Table 1 shows Cohn-Kanade database specifications.

TABLE I
COHN-KANADE DATABASE

Age	18 to 50
Female	69%
Male	31%
Euro-American	81%
Afro-American	13%
Other groups	6%
Resolution	640×490 Grayscale

III. Facial Feature Point Tracking Using Optical Flow

In the first digitized frame, 21 key feature points were manually marked with a computer-mouse around facial features such as eyes, eyebrows and mouth (Fig. 1). Among 44 AUs in the Facial Action Coding System (FACS), 39 AUs are directly associated with the movement of eyes, eyebrows and mouth. That is why the information expressing movement of eyes, eyebrows and mouth is desirable for machine recognition of the facial expressions. We are confined in these three components and then determine the facial feature points which are representative of the boundary between these components and skin. Some of this points used only for specifying the model features in the first frame. Other points were automatically tracked in the subsequent frames using cross-correlation based optical flow.

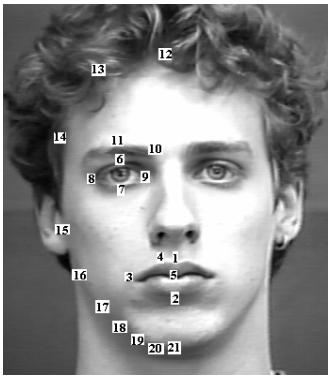


Fig. 1 Selected 21 facial feature points

Each point is the center of a 11×11 flow window that includes horizontal and vertical flows. Fig. 2 shows the implementation of this method in two subsequent frames:

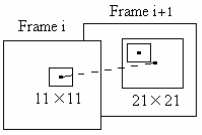


Fig. 2 Cross-correlation optical flow calculation

Cross-correlation of 11×11 window in the first frame, and a 21×21 window at the next frame were calculated and the position with maximum cross-correlation of two windows, were estimated as the position of the feature point at the next frame.

In our earlier researches, we extracted some features from normalized position of the feature points in the first frame and the last frame (The position of all feature points was normalized by position of the tip of nose). This features used to analyzing facial expressions in to one of the six primary emotions or into a set of single or composite AUs. For classifying the facial expressions we used the fuzzy logic and the RBF neural networks [7-10].

In this paper we developed a face model for synthesizing the facial expressions, by tracking some feature points that were described in the Fig. 1. In the subsequent section we describe the specifications of the proposed face model.

IV. Face Model

We are confined in the three facial features: mouth, eyes and eyebrows. We used a muscle-based triangular patch object model that the vertices of the triangles were determined from the feature points tracking results. Deformed triangles gave the ability of shape tracking to the model. In this section we describe the mouth, eyes and eyebrows models.

A. Mouth Model

Fig. 3 shows the selected mouth feature points and the features that were extracted from these points.

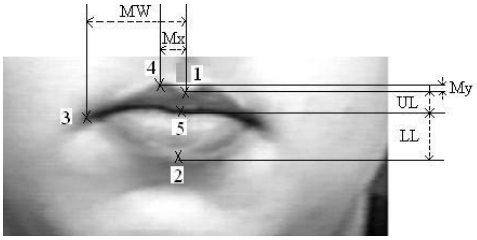


Fig. 3 Mouth features and feature points

Points 1, 2 and 3 were tracked in the subsequent frames and theirs position were transferred to the mouth model. But other points were only used in the first frame for determining some features such as mouth wide, upper and lower lip thickness.
Fig. 4 shows the proposed mouth model. The coordinates of the vertices 1, 8 and 15 directly determined from the normalized position of the 1, 2 and 3 feature points in the Fig. 3 (UL were normalized to one). Coordinates of the other vertices were determined from the features MW, Mx, My, UL and LL relative to 1, 8 and 15 vertices. Right hand side of the model is the symmetry of the left hand side.

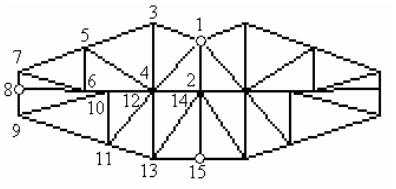


Fig. 4 Mouth model

B. Eye and Eyebrow Model

Fig. 5 shows the selected eye and eyebrow feature points and the features that were extracted from these points.

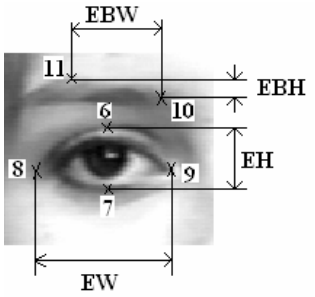


Fig. 5 Eye and Eyebrow features and feature points

Points 6, 7, 10 and 11 were tracked in the subsequent frames and the other points were only used in the first frame for determining some features such as wide and height of the eye and eyebrow.
Fig. 6 shows the proposed eye and eyebrow model. The coordinates of the vertices 16, 17, 30 and 31 were directly determined from normalized position of the 6, 7, 10 and 11 feature points in the Fig. 5. Coordinates of the other vertices were determined from features EH, EW, EBH and EBW relative to 16, 17, 30 and 31 vertices. For example wide of the iris assumed to be a half of the EW. Right hand side of the model is the symmetry of the left hand side.

Face landmarks were determined by using 10 landmark points around the face in the first frame (points 12-21 in the Fig. 1). These points weren't tracked in subsequent frames, but the points around the jaw were synchronized with the movement of the point 2 in the Fig. 3.

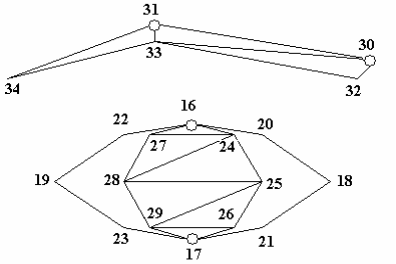


Fig. 6 Eye and Eyebrow model

V. Experimental Results

The proposed face model was evaluated with the Cohn-Kanade database and our prepared database. Fig. 7 shows the result of feature points tracking in the original face and the model deformation to show the same expressions. Tracked feature points in the original frames were also denoted in this figure.

The proposed model also can be used for lip tracking and speech synchronizing facial animation system. Fig. 8 shows the results of lip tracking for pronouncing the word "Salam", that means "Hello" in Persian.

VI. Conclusion

In this paper we developed a deformable muscle-based face model that tracks some FCP's in the real face image sequences and shows the same expressions. The proposed deformable model had a simple structure and used a few set of control points comparing to the similar face models. The proposed model has a low complexity and is suitable for real time implementations, such as real time facial animation. Because of using the frontal images, we used a 2-D face model. By using the side images beside the frontal images and applying 3-D coordinates to the vertices of the model, 3-D face model can be obtained.

Acknowledgment

The authors appreciate the support of Iranian Telecommunication Research Center (ITRC) for this research and would like to thank J.F. Cohn and T. Kanade from Pittsburgh University for their kindly providing the image database.

References

[1] P. Ekman and W.V. Friesen, Facial Action Coding System (FACS), *Consulting Psychologists Press*, Inc., 1978.
[2] Z. Wu,P. S. Aleksic and A. K. Katsaggelos," Lip Tracking for MPEG-4 Facial Animation," in *Proc. 4th IEEE Int. Conf. on Multimodal Interfaces,ICMI'02,*
[3] P. Eisert, T. Wiegand and B. G. Fellow," Model-Aided Coding: A New Approach to Facial Animation into Motion Compensated Video Coding," *IEEE Trans. On Circuits and Systems for Video Technology, vol. 10, No. 3, April 2000.*

[4] S. T. Worral, A. H. Sadka, and A. M. Kondoz, "3-D Facial Animationfor Very Low Bit Rate Mobile Video," in *Proc. IEE Int. Conf. on* 3G Mobile Communication Technologies, May 2002.
[5] F. Erol, "Modeling and Animating Personalized Faces," *M.Sc. Thesis*, Bilkent university, January 2002.
[6] T. Kanade, J. Cohn and Y. Tian. *Comprehensive database for facial expression analysis*, 2000.
[7] H. Seyedarabi, A. Aghagolzadeh and S. Khanmohammadi, "Facial Expressions Recognition from Static Images using Neural Networks and Fuzzy logic," *The 2nd Iranian Conference on Machine Vision and Image processing (MVIP2003)*,vol.1 pp 7-12, Tehran, 2003.
[8] H. Seyedarabi, A. Aghagolzadeh and S. Khanmohammadi, "Facial Expressions Recognition from Image Sequences using Cross-correlation Based Optical-Flow and Radial Basis Neural Networks," *The 12th Iranian Conference on Electrical Engineering,(ICEE 2004),* vol.1 pp 165-170,Mashhad, 2004 .
[9] H. Seyedarabi, A. Aghagolzadeh and S. Khanmohammadi, "Recognition of Six Basic Facial Expressions by Feature-Points Tracking using RBF Neural Networks and Fuzzy Inference System," *The IEEE International conference on Multimedia & Expo (ICME2004),* Taipei, Taiwan ,June 2004.
[10] Ali Aghagolzadeh, Hadi Seyedarabi and Sohrab Khanmohammadi ," Single and Composite Action Units Classification in Facial Expressions by Feature-Points Tracking and RBF Neural Networks", *Ukrainian Int. conf. on signal/Image processing,UkrObraz 2004*, October 2004, Kiev, Ukraine.

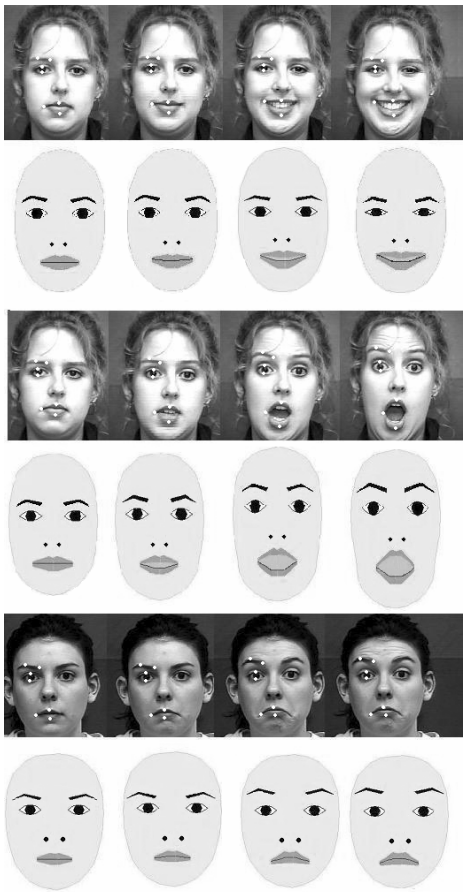


Fig. 7 Facial expressions tracking and animating

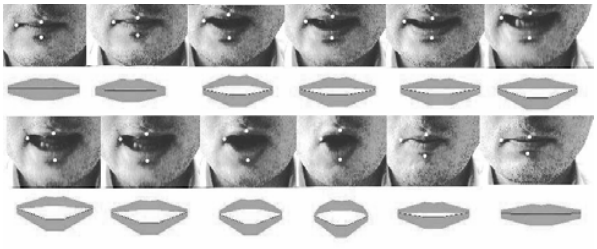
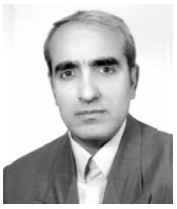


Fig. 8 Lip tracking for pronouncing the word "Salam"

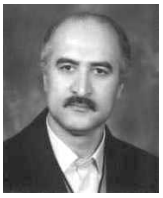
Hadi Seyedarabi Received B.S. Degree in Electrical engineering from University of Tabriz, Iran, in 1993 and the M.S. degree in Telecommunication engineering from K.N.T. University of technology, Tehran, Iran in 1996.

He is currently a PhD student in Electrical engineering at the University of Tabriz. His research interests are image processing, computer vision and facial animation.



Ali Aghagolzadeh received B.S. degree from University of Tabriz, Tabriz, Iran, in 1985 and M.Sc. degree from Illinois Institute of Technology, Chicago, IL, USA, in 1988 and Ph.D. degree from Purdue University, west Lafayette, IN, USA, in 1991 all in Electrical Engineering.

He is currently an associate professor of Faculty of Electrical and Computer Engineering in University of Tabriz, Tabriz, Iran. His research interests are Digital Signal Processing, Digital Image Processing, Computer Vision and Human - Computer Interaction



Sohrab Khanmohammadi Received the B.S. Degree in Industrial Engineering from the Sharif University, Tehran, Iran, in 1976. M.S. degree in Automatic Engineering, Paul Saba tier University, France, 1980. Special Diploma in Advanced Automatics and Systems from Ecole Nationale Superieure de l'Aeronautiqe et de l'Espace, France, 1981; and Ph.D. (Doctor Engineer) in Automatic Engineering from Ecole Nationale Superieure de l'Aeronautiqe et de l'Espace, France, 1983.

He is a Professor of Control and System in Faculty of Electrical Engineering, Tabriz University. He is a lecturer member of Sharif University and I. A. U Science and Research Campus in Tehran. He was also the professor of fuzzy systems in University of Western Ontario during his sabbatical in Canada.